RSAC 2025 Conference

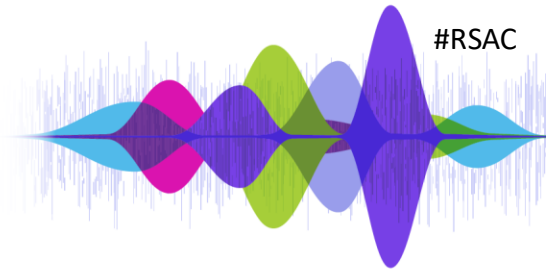
Many Voices. One Community.

SESSION ID: HTA-R02

# Ironcat Unleashed: Outwitting AI and Antivirus with Data Driven Malware

**Aaron Rosenmund**

Managing Director of Tradecraft & Programs
OnDefend
www.linkedin.com/in/aaronrosenmund/



# Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the presenters individually and, unless expressly stated to the contrary, are not the opinion or position of RSA Conference LLC or any other co-sponsors. RSA Conference LLC does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented.

Attendees should note that sessions may be audio- or video-recorded and may be published in various media, including print, audio and video formats without further notice. The presentation template and any media capture are subject to copyright protection.

© 2025 RSA Conference LLC or its affiliates. The RSAC and RSAC CONFERENCE logos and other trademarks are proprietary. All rights reserved.

# The Agenda

- Current State Of AI Malware
- Designing True AI Generated Malware
- Building, and Running AI Malware
- Where Does This Path Lead?
- What's Next For You?

RSAC 2025 Conference

# Current State of AI Malware

# What Everyone Thought Would Happen

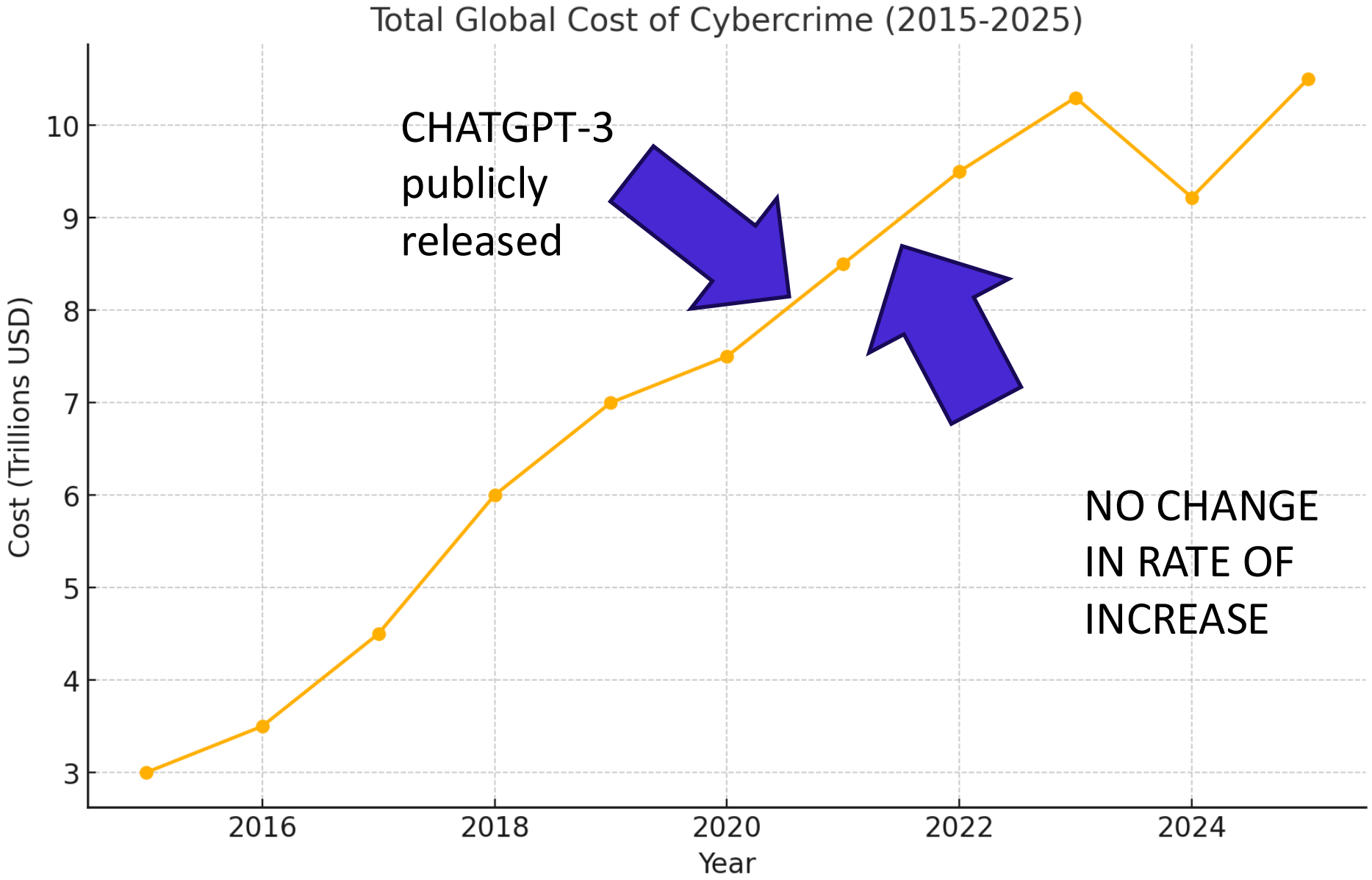
# What Actually Happened

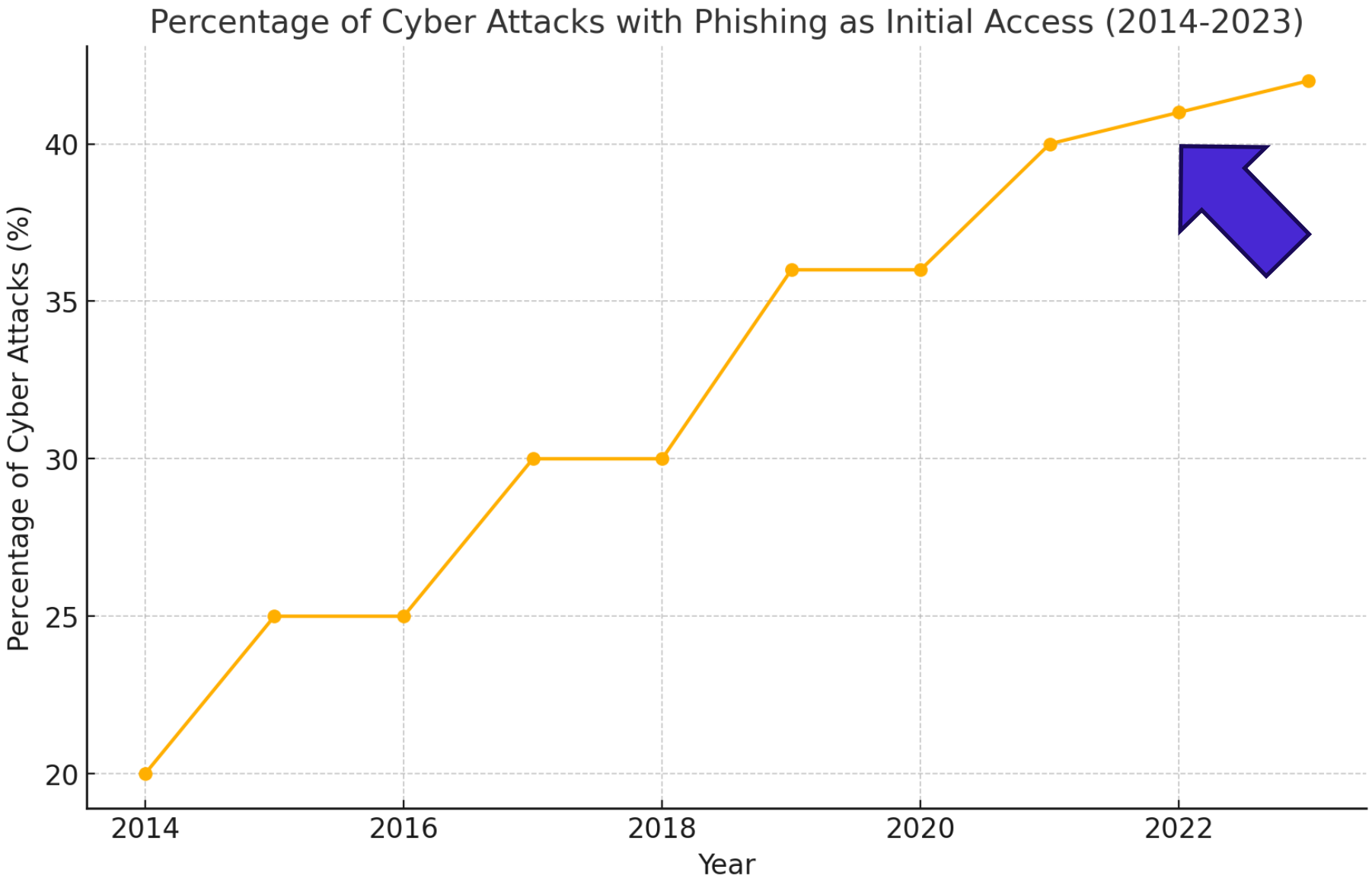
# AI's Current Role In Malware

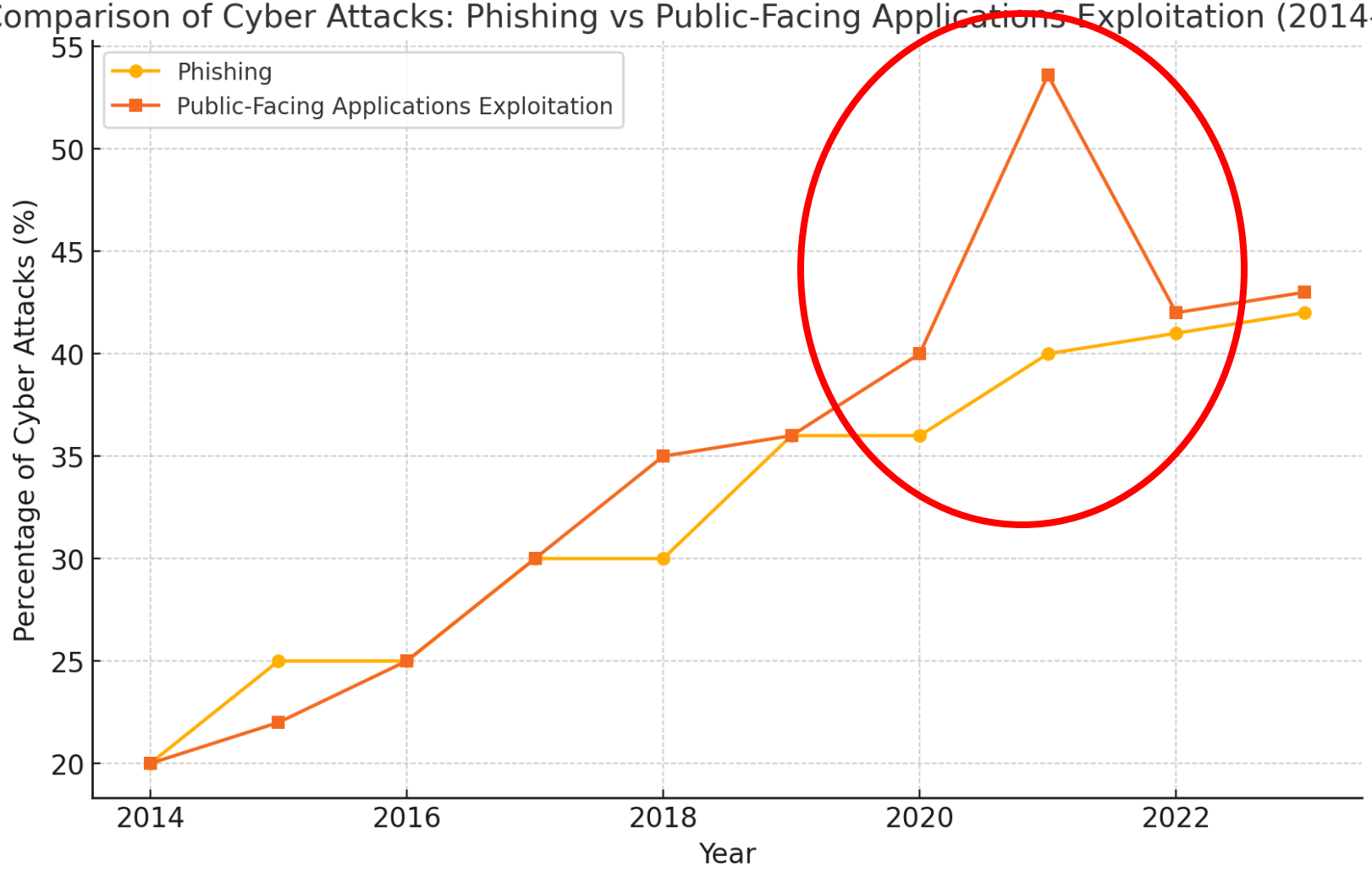
## What has been observed

## AI – Supported Malware

- AI-driven phishing attacks
- Deepfakes
- Adversarial AI/ML
- Malicious GPTs
- Ransomware attacks (optimized)


Total Global Cost of Cybercrime (2015-2025)
CHATGPT-3 publicly released
NO CHANGE IN RATE OF INCREASE
Cost (Trillions USD)
Year
3
4
5
6
7
8
9
10
2016
2018
2020
2022
2024


Percentage of Cyber Attacks with Phishing as Initial Access (2014-2023)
Percentage of Cyber Attacks (%)
20
25
30
35
40
2014
2016
2018
2020
2022
Year


Comparison of Cyber Attacks: Phishing vs Public-Facing Applications Exploitation (2014-2023)
Phishing
Public-Facing Applications Exploitation
Percentage of Cyber Attacks (%)
55
50
45
40
35
30
25
20
2014
2016
2018
2020
2022
Year

RSAC 2025 Conference

# Current State of Malware Detections

## Anti-Virus and Endpoint Detection & Response

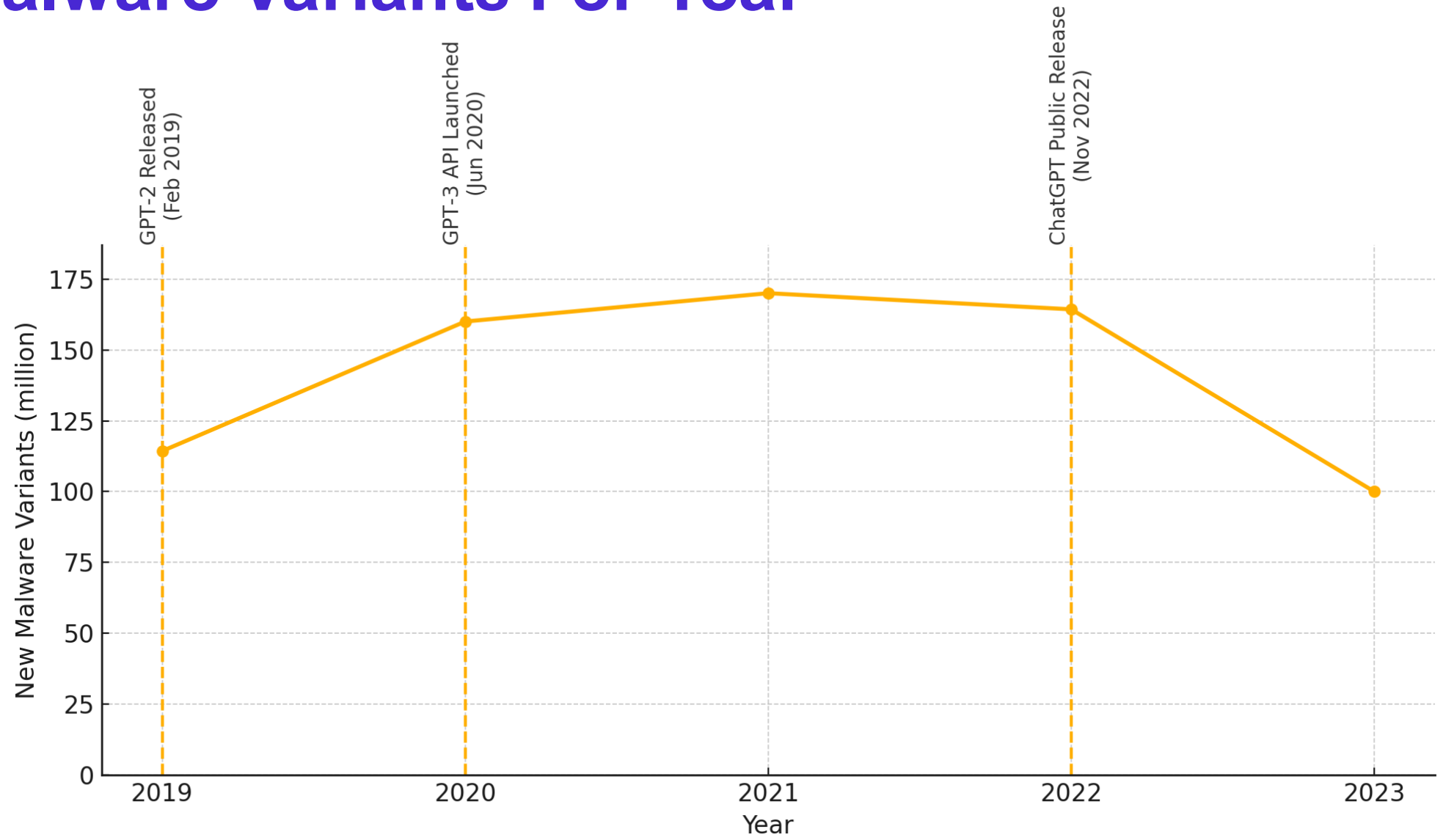
Many Voices.
**One Community.**

# AI's Current Role In Detection Technology

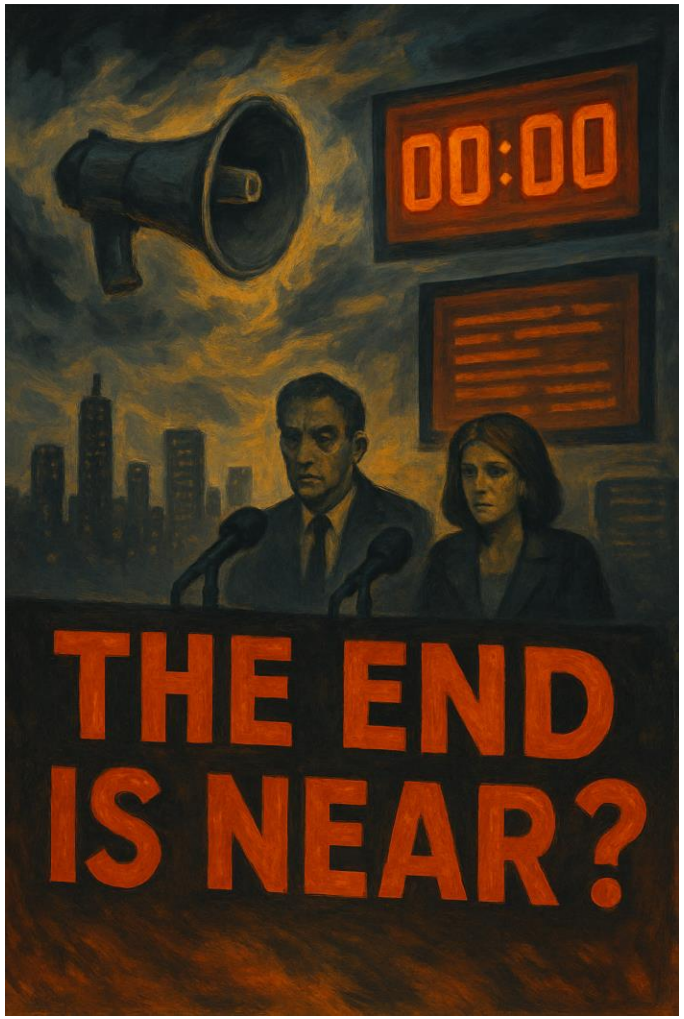
## AI – Supported Anti-Virus

- Cloud Powered Threat Intel
- Real-Time Program-Flow Analysis
- Deep Learning At Scale
- Behavioral ML Models
- Widespread Adoption of AI for Defense

# Malware Variants Per Year



# Despite the Rhetoric



# UNTIL NOW

RSAC 2025 Conference

# Designing True AI Generated Malware

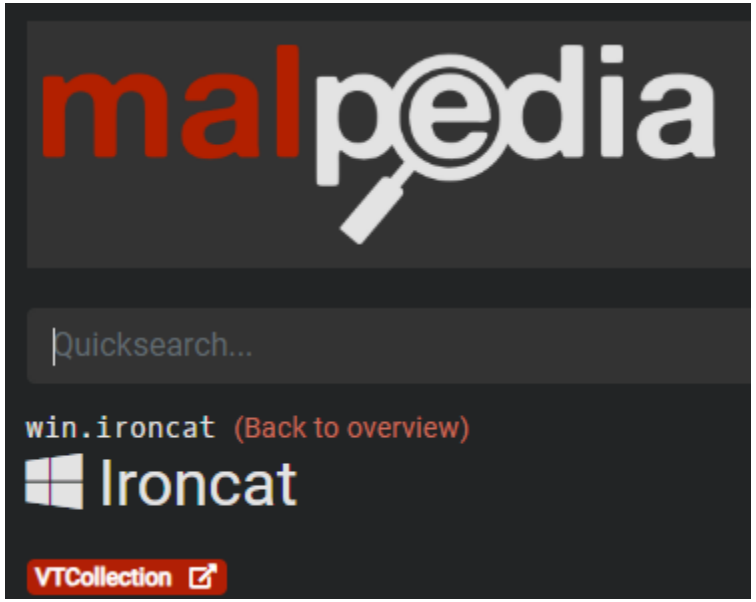
## Like Most AI Solutions, It Is Mostly Data Science

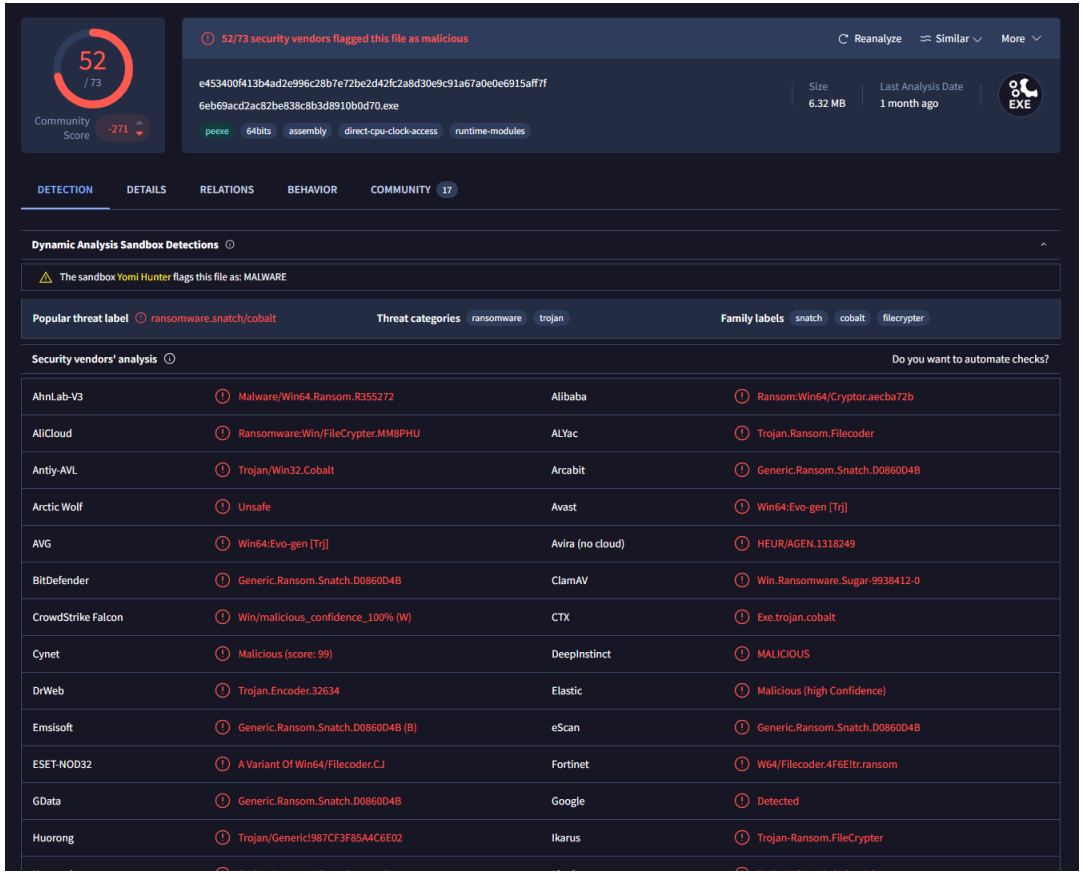
Many Voices.
**One Community.**

# Start With Something Known

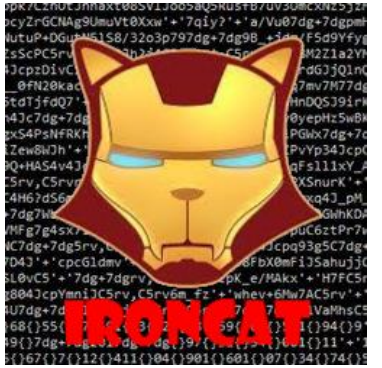
Ransom:Win32/Ironcat.GD!MTB

Detected by Microsoft Defender Antivirus

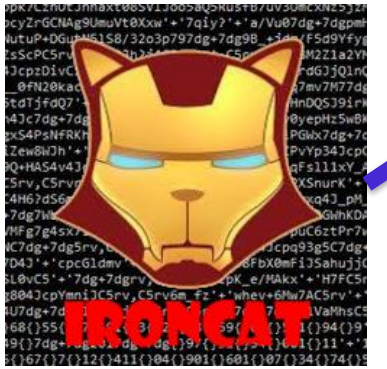


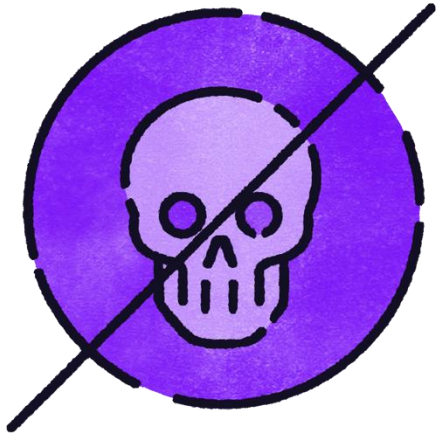
# IRONCAT MALWARE



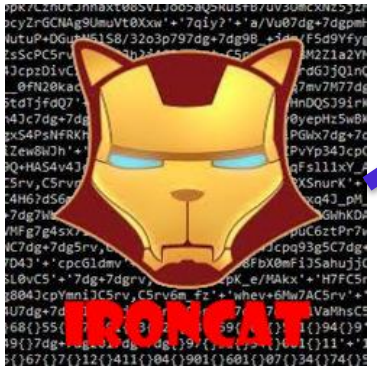
IRONCAT MALWARE



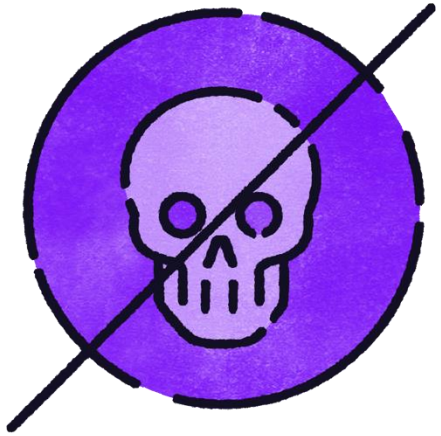

ANTI-VIRUS CHECK



IRONCAT MALWARE

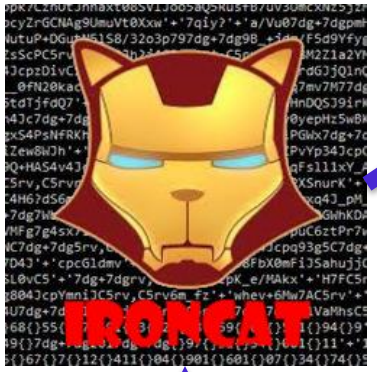



ANTI-VIRUS CHECK

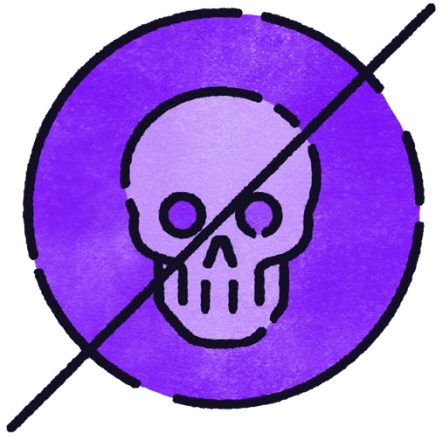


DETECTION

IRONCAT MALWARE

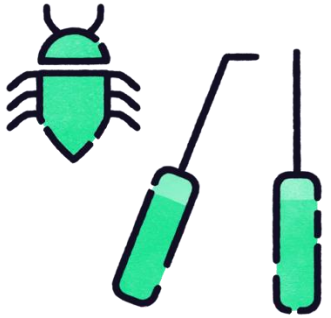



ANTI-VIRUS CHECK



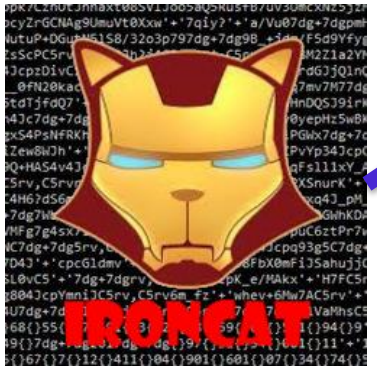
DETECTION

ANALYZE DETECTION STRINGS

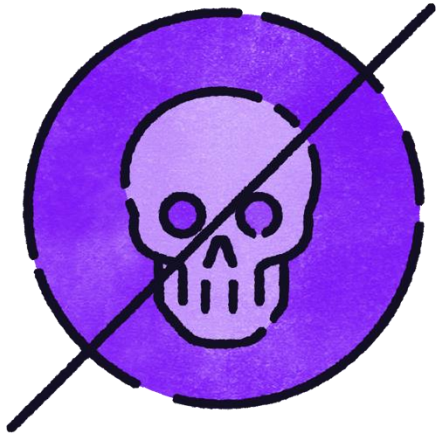


MODIFY MALWARE SOUCE CODE OR BINARY

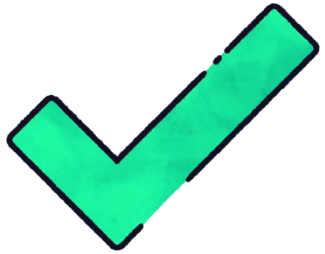
NEW IRONCAT MALWARE



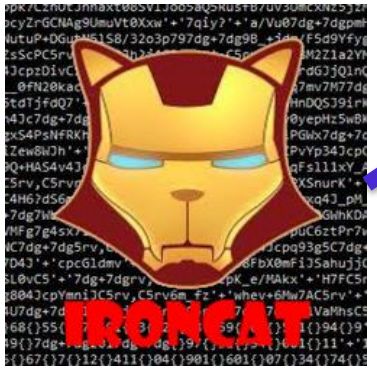

ANTI-VIRUS CHECK



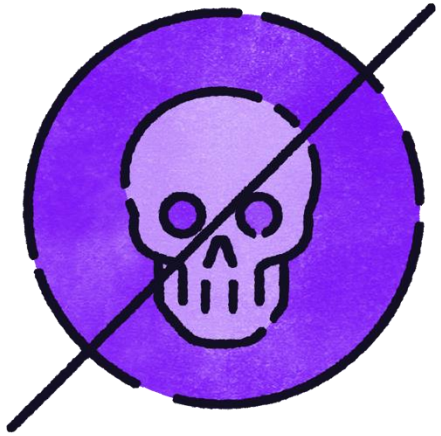
NO DETECTIONS



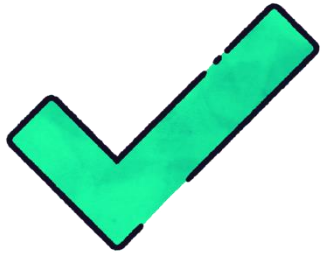
NEW IRONCAT MALWARE



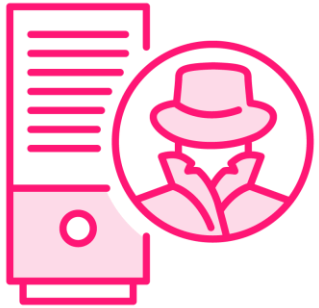

ANTI-VIRUS CHECK



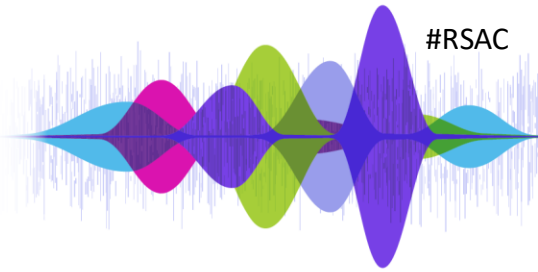
NO DETECTIONS



INFECT TARGET



# Two Sets of Rules to Evade



## File Based Detections

- Malicious Strings
- Yara Heuristics for Assembly Behavior
- File size restrictions
- File Hashes
- Fuzzy Hashes

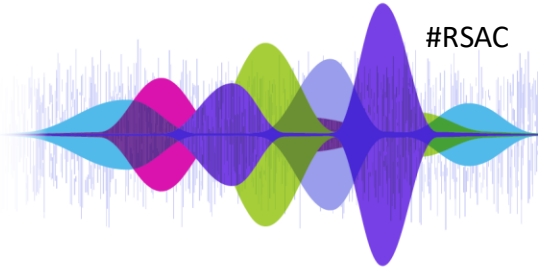
## In Memory Detections

- Still mostly string based
- Post binary unpacking
- API call monitoring

# Manual Evasion Requires A Lot of Work

# So Offload that Work to AI!

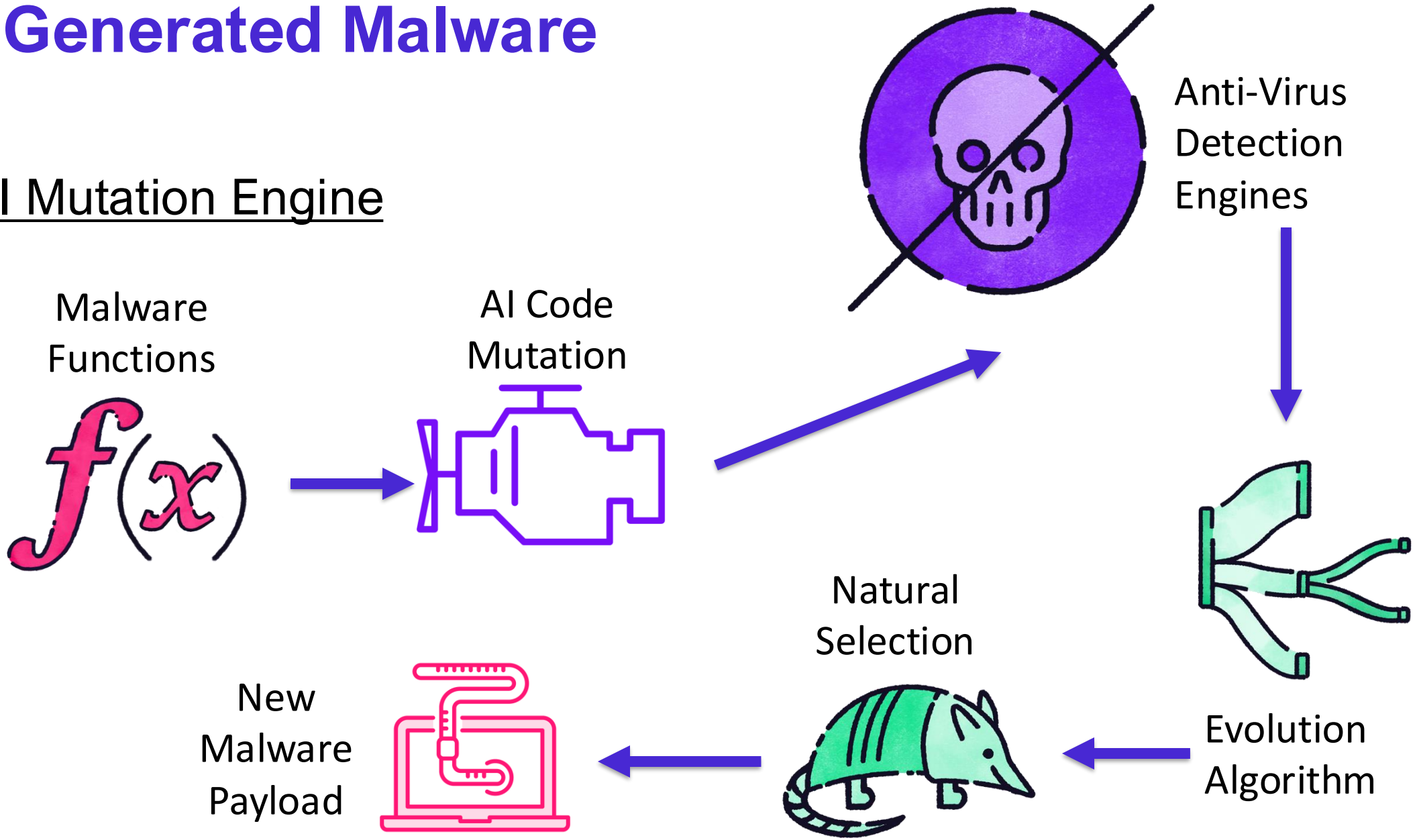
# Two Kinds of AI Involvement in Malware



- AI Generated Malware
  - Polymorphic
  - Learns from detections
  - Speed of operation
- Malware with Built In AI (Intelligent Malware)
  - Autonomous operation
  - Speed of compromise

# AI Generated Malware

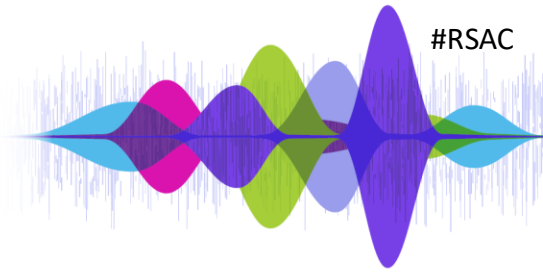
- AI Mutation Engine



# AI Generated Malware

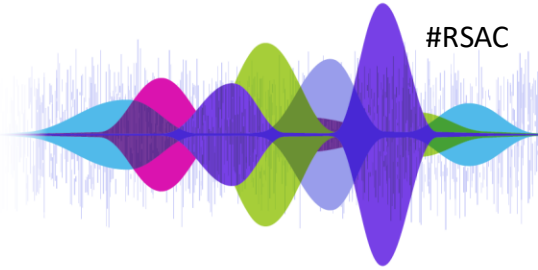
- Desired Function
  - C2 Communication
  - Encryption
  - New Process Spawning(Injection)
  - Enumeration of Target System
  - Web Shell

# Controlled Chaos Framework



- Vary how a program behaves, using as set of functions or actions but ensuring that the outcomes are creatively diverse with logical constraints. While steering the creativity of the system with input of the Anti-Virus detections.

# A Model for Creativity in Variations



- Markov Chains
  - Favor “natural” paths based on successful previous runs.
- Genetic Algorithms
  - Keep top 10% undetected payloads, mutate the others (natural selection)
- Neural Guidance
  - Feedback metadata of detections into a LLM to suggest next best mutation strategies

# Natures Way

## Mutation of Functions (Traits)

```
🔁 Mutation round 1 of 3...
📄 Reading file: .\source-code\functions\encryption\encryption-001.go
🧠 Mutation level: light
✅ GPT response received.
🔍 Reviewing for compilation issues...
💾 Writing output to: .\source-code\functions\encryption\encryption-002.go
✅ Round 1 complete in 82.49 seconds.

🔁 Mutation round 2 of 3...
📄 Reading file: .\source-code\functions\encryption\encryption-002.go
🧠 Mutation level: light
✅ GPT response received.
🔍 Reviewing for compilation issues...
💾 Writing output to: .\source-code\functions\encryption\encryption-003.go
✅ Round 2 complete in 88.88 seconds.

🔁 Mutation round 3 of 3...
📄 Reading file: .\source-code\functions\encryption\encryption-003.go
🧠 Mutation level: light
✅ GPT response received.
🔍 Reviewing for compilation issues...
💾 Writing output to: .\source-code\functions\encryption\encryption-004.go
✅ Round 3 complete in 90.10 seconds.
PS C:\ironcat-unleashed\ironcat-unleashed>
```

## Evolution of Species

```
[=] Files processed: 12
[=] Snippet stats collected:
    - encryption-003.go: {'uses': 1, 'detections': 7}
    - webshell-002.go: {'uses': 1, 'detections': 1}
    - webshell-006.go: {'uses': 1, 'detections': 1}
    - webshell-003.go: {'uses': 1, 'detections': 1}
    - webshell-001.go: {'uses': 2, 'detections': 2}
    - encryption-002.go: {'uses': 1, 'detections': 7}
    - encryption-004.go: {'uses': 2, 'detections': 15}
    - encryption-001.go: {'uses': 1, 'detections': 8}
    - webshell-004.go: {'uses': 1, 'detections': 1}
    - webshell-005.go: {'uses': 1, 'detections': 1}
IRONCAT EVOLUTION COMPLETE
```

RSAC 2025 Conference

# Building & Running AI Malware

RSAC 2025 Conference

# DEMO TIME

## Live Demo of Operational Mutation Engine

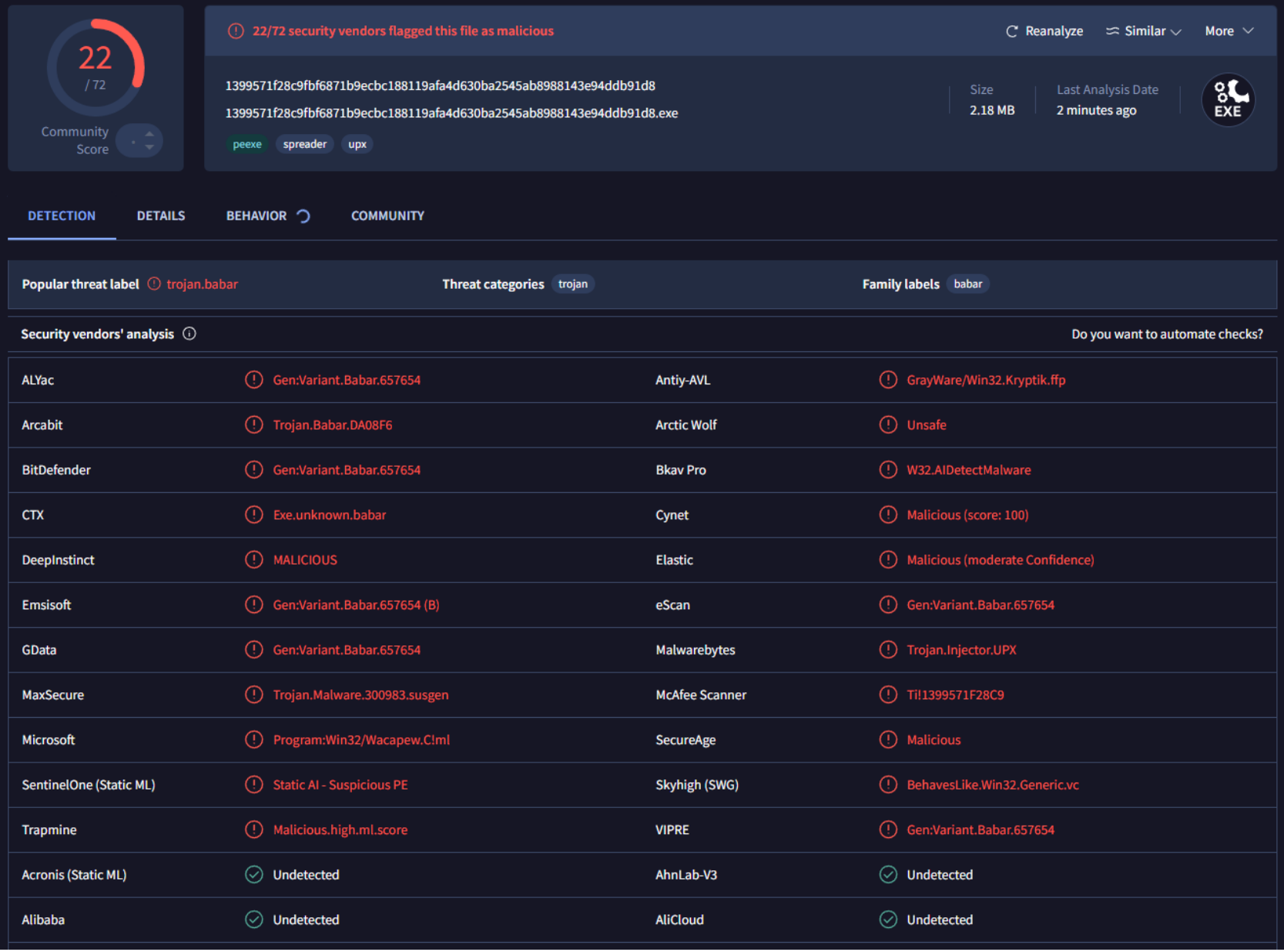
Many Voices.
**One Community.**

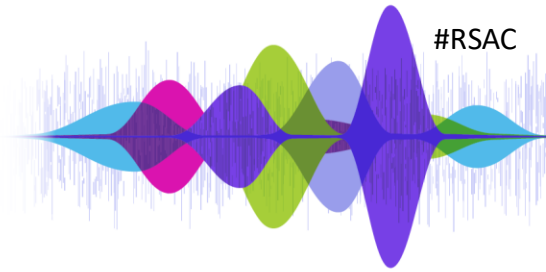
# A Note About UPX

```
unleashed\ironcat-unleashed\3-gen-payload> python .\gen-payload.py --input ..\1-source-code\ic-v1.go --output .\gen-samples\ --
RCH=386 CGO_ENABLED=0 garble -tiny -literals build -ldflags -s -w -trimpath -o D:\1-ironcat-data\ironcat-unleashed\ironcat-unle

data\ironcat-unleashed\ironcat-unleashed\3-gen-payload\gen-samples\temp_build.exe with UPX...
te Packer for eXecutables
yright (C) 1996 - 2025
mer, Laszlo Molnar & John Reiser   Feb 20th 2025

io      Format      Name
---   -----------   -----------
34%    win32/pe     temp_build.exe


cat-data\ironcat-unleashed\ironcat-unleashed\3-gen-payload\gen-samples\1399571f28c9fbf6871b9ecbc188119afa4d630ba2545ab8988143e
oncat-data\ironcat-unleashed\ironcat-unleashed\3-gen-payload\gen-samples\1399571f28c9fbf6871b9ecbc188119afa4d630ba2545ab898814
unleashed\ironcat-unleashed\3-gen-payload>
```

## UPX IS A COMMON PACKING TOOL FOR MALWARE

But, it is not random, and does leave artifacts.

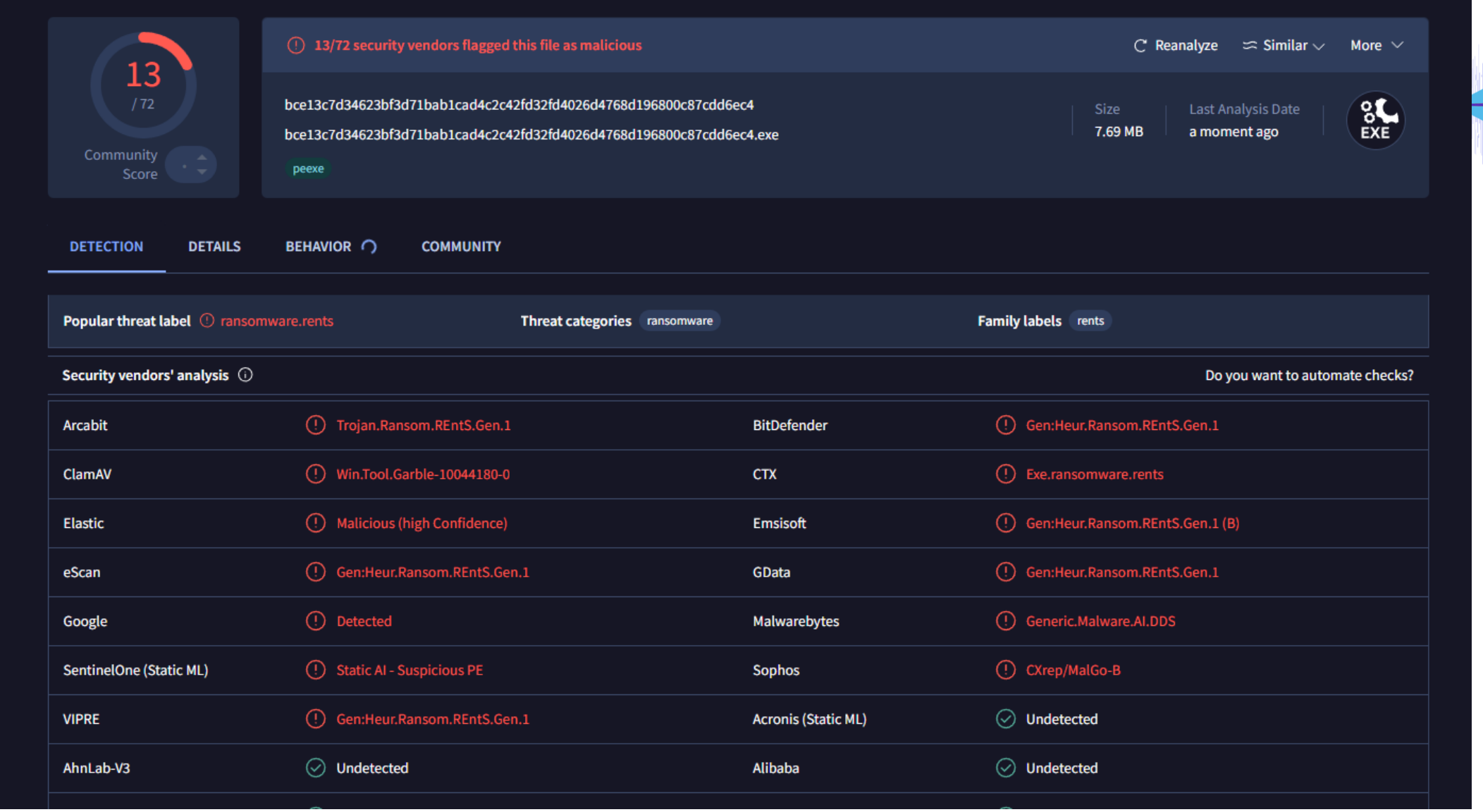



# UPX Flagged Binary

OnDefend

RSAC 2025 Conference



13 / 72

Community Score

13/72 security vendors flagged this file as malicious

Reanalyze | Similar | More

bce13c7d34623bf3d71bab1cad4c2c42fd32fd4026d4768d196800c87cdd6ec4

bce13c7d34623bf3d71bab1cad4c2c42fd32fd4026d4768d196800c87cdd6ec4.exe

peexe

Size 7.69 MB

Last Analysis Date a moment ago

EXE

DETECTION | DETAILS | BEHAVIOR | COMMUNITY

Popular threat label: ransomware.rents

Threat categories: ransomware

Family labels: rents

Security vendors' analysis

Do you want to automate checks?

| | | | |
|---|---|---|---|
| Arcabit | Trojan.Ransom.REntS.Gen.1 | BitDefender | Gen:Heur.Ransom.REntS.Gen.1 |
| ClamAV | Win.Tool.Garble-10044180-0 | CTX | Exe.ransomware.rents |
| Elastic | Malicious (high Confidence) | Emsisoft | Gen:Heur.Ransom.REntS.Gen.1 (B) |
| eScan | Gen:Heur.Ransom.REntS.Gen.1 | GData | Gen:Heur.Ransom.REntS.Gen.1 |
| Google | Detected | Malwarebytes | Generic.Malware.AI.DDS |
| SentinelOne (Static ML) | Static AI - Suspicious PE | Sophos | CXrep/MalGo-B |
| VIPRE | Gen:Heur.Ransom.REntS.Gen.1 | Acronis (Static ML) | Undetected |
| AhnLab-V3 | Undetected | Alibaba | Undetected |



Same Binary
Without UPX

# Garble Has The Same Result

```
Analysis results saved to ./3-generated-vtresults\cfca6b60a6a5e9fbf9dd6e831b4be88ccdde

✓ Submission succeeded!
VirusTotal Detection Stats:
          Malicious: 12
         Suspicious: 0
         Undetected: 60
           Harmless: 0
            Timeout: 0
  Confirmed-timeout: 0
            Failure: 0
   Type-unsupported: 4

Detected by the following engines:
Elastic: malicious (high confidence)
MicroWorld-eScan: Gen:Variant.Midie.163825
ClamAV: Win.Tool.Garble-10044180-0
BitDefender: Gen:Variant.Midie.163825
Emsisoft: Gen:Variant.Midie.163825 (B)
VIPRE: Gen:Variant.Midie.163825
CTX: exe.unknown.midie
Sophos: CXrep/MalGo-B
GData: Gen:Variant.Midie.163825
Arcabit: Trojan.Midie.D27FF1
Google: Detected
ALYac: Gen:Variant.Midie.163825
Exiting after successful deployment.
```

- Use of Garble is also signatured pretty heavily.

# Combined Functions Lowered Scores

```
Analysis results saved to ./3-generat

✓ Submission succeeded!
VirusTotal Detection Stats:
           Malicious: 3
          Suspicious: 0
          Undetected: 69
            Harmless: 0
             Timeout: 0
   Confirmed-timeout: 0
             Failure: 0
    Type-unsupported: 4

Detected by the following engines:
Ikarus: Trojan-Dropper.WinGo.Agent
Webroot: W32.Trojan.Gen
Google: Detected
Exiting after successful deployment
```

- Same options with neither UPX or Garble
  - Encryption
  - Web Shell

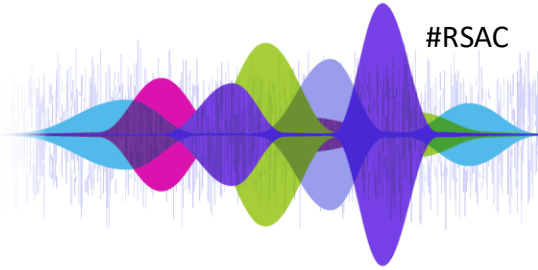
RSAC 2025 Conference

# Where Does This Path Lead?

## The Future of AI Malware

Many Voices.
**One Community.**

# Where Does This Path Lead?



## AI Mutation Engines

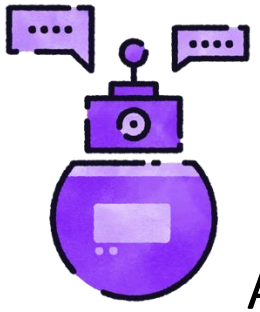
- Malware Devs will write generators, not bother with a specific sample for a specific target.
- AI becomes a must have capability in malware dev moving forward
- Legacy malware will quickly becomes obsolete (could be good) raises the bar
- New defenses required

## AI C2 Systems

- This is wargames
- Speed of compromise will significantly increase
- Defense in depth is significantly weakened
- Anti-Virus/EDR by nature can't keep up.
- As A Service Model

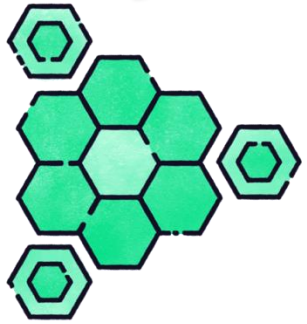
# AI IN the Malware

- Smart Malware
  - Modular
  - LLM Runs on C2 Server



AI
AGENT

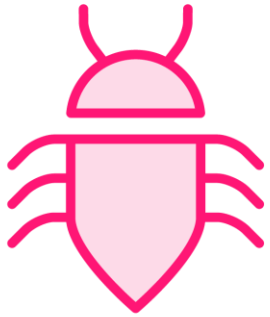
Target VectorDB



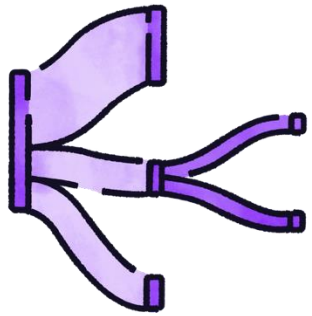
Reasoning LLM

C2 Server
Controller

Payload
Enumeration



Deliver
Module for
Next Steps
over C2

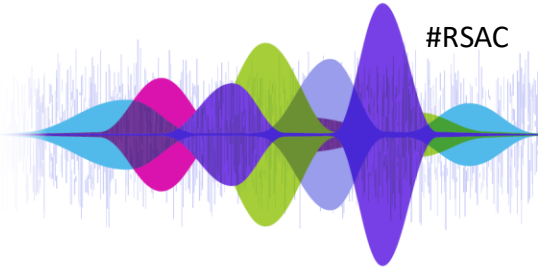


AI Mutation
Engine -
Modules

# AI IN the Malware

- AI System resides on C2 Network
- Enumeration and environment send situational awareness information to C2 server
- C2 Sever logic sends information to AI agent to identify best path forward
- Agent then asks AI mutation engine for a polymorphic payload module for selected steps
- Process repeats

# Apply What You Have Learned



- Next week you should:
  - Use the mutation engine to check your own security systems in on an isolated system.
- In the first three months following this presentation you should:
  - Evaluate current defenses that over focus on anti-virus and EDR detections.
- Within six months you should:
  - Design future cybersecurity strategy to strengthen against AI threat, and TEST YOUR DEFENSES!

RSAC™ 2025 Conference

Many Voices.
**One Community.**

**Github.com/arosenmund/ironcat-unleashed**